Human Understandings of Probability

Michael Smithson, The Australian National University

"The deep study of certain games will perhaps lead to a new chapter of the theory of probabilities... It will be a new science where psychology will be no less useful than mathematics, but this new chapter will be added to previous theories without modifying them." Borel, E. (1924/1964: 58-59).

Probabilities as Degrees of Belief

The concept of a probability as reflecting a degree of belief is the principal connection between probability theories and cognitive psychology. It is all too easy to forget that the concept of probability is historically and culturally specific, and its connection with psychological uncertainty even more so. There still exist cultures with no identifiable notion of probability, and some people in Western cultures disavow any connection between probability and mental aspects of uncertainty such as degrees of belief. This chapter therefore begins with a survey of the development of this connection.

The earliest scholar to link probability with degree of belief most likely is Jacob Bernoulli in his 1713 *Ars Conjectandi*, with Laplace, De Morgan, and Donkin elaborating this view during the first half of the 19th century. Keynes (1921) is noted for having added that subjective probability judgments are logical relations between one set of propositions and another, conditioned in some sense by the knowledge available to the judge. In making this connection, Keynes could account for two reasonable judges assigning different probabilities to the same prospect by differing in the information available to them. Keynes also famously declared that not all logical probabilities are quantifiable or even comparable with one another. As Ramsey (1926/1931) observed, this claim created difficulties for the one-to-one correspondence between degrees of belief and degrees of probability relations.

Borel (1924/1964) and Ramsey disputed Keynes on this latter point and related matters. Ramsey began by arguing that there is no apparent distinction between so-called “quantifiable” and “unquantifiable” beliefs, although he allowed that some beliefs can be measured more accurately and/or reliably than others. Importantly, he dismissed introspection as a valid and accurate source of such measurements and turned, instead, to what is now the standard behavioral approach based on betting-rates, generalizing it to an account based on preferences. He thereby claimed to have found a “purely psychological” way of measuring degrees of belief.

It is at this point that Ramsey raised the question of what constitutes “reasonable” degrees of belief. From then on, concepts of coherence were elaborated, for example, by Ramsey and de Finetti (1937/1964), either in the sense of avoiding sure loss or a stronger sense than that. Subjective probability theories became more prescriptive and less descriptive. Neoclassical economics sprang directly from these developments. A relatively independent contemporaneous contribution came from the invention of game theory, again initiated by Borel (1921/1953) and then famously elaborated by von Neumann and Morgenstern (1944). Nevertheless, it was not until the 1950’s that psychologists (principally Ward Edwards, 1954 & 1961, and John Cohen, 1954 & 1960) saw a need for a non-prescriptive, psychologically informed, understanding of human probability judgments.

The nature of the boundary between the prescriptive and descriptive turned out to be contestable. Cohen (1960: 28-29) wrote to de Finetti expressing the view that the psychological study of probability does not involve attributions of "error" to humans. Cohen quoted de Finetti's response, which begins "Unlike almost all mathematicians, I agree completely with your statement that every probability evaluation is a probability evaluation, that is, something to which it is meaningless to apply such attributes as right, wrong, rational, etc." Immediately thereafter, however, de Finetti makes it clear that he regards incoherent bets as "nonsensical" by which he appeared to mean irrational.

Nonetheless, it still seemed possible that human probability judgments might correspond to the laws of probability in some respects, perhaps along the lines of the Weber-Fechner laws of human perceptions of physical properties. Indeed, a tradition of modeling subjective probabilities via continuous weighting functions survives to this day. However, studies by Cohen and his colleagues in the 1950's and 1960's (summarized in Cohen, 1964) of probability estimation and reasoning in children and adults suggested that even adult judgments not only are miscalibrated but also deviate substantially from probability theory. Several of their findings anticipated later research, such as overestimation of low-probability events, intransitive choices among gambles, and the conjunction fallacy (assignment of higher probability to the conjunction of two events than to either constituent event).

A decade later, Amos Tversky and Daniel Kahneman began publishing research that became known as the "heuristics and biases" school, highlighting what they portrayed as human cognitive illusions and errors in probability judgments (e.g., Tversky & Kahneman 1974, Kahneman, Slovic, & Tversky 1982). The literature on this topic is large, and so is the list of "errors" (Hogarth 1980 reviewed 27). We shall examine several of them in sections to come. Some researchers underpinned claims that heuristics and biases are genuinely irrational by identifying correlations between scores on tests of mental abilities and the tendency to use normative strategies in judgment and decision making (Stanovich & West 2000). Cohen, Edwards, and Herbert Simon disagreed (for somewhat different reasons) with the heuristics and biases school's emphasis on human errors. During the 1980's and 1990's a variety of debates ensued about whether particular human deviations from standard probability theoretic norms are irrational or not.

At about the same time, Simon (1956, 1982) elaborated his "bounded rationality" framework. He argued that human judgment and reasoning are limited by their bounds on cognitive capacity, the time available to gather and process information, and the information available. Humans therefore adopt "satisficing" solutions to problems instead of optimal ones whose requirements on cognitive capacity, time, and information may exceed those limits. Unlike exponents of the heuristics and biases school, Simon did not regard satisficing strategies as irrational or erroneous, but instead adaptive.

Gigerenzer and his colleagues expanded Simon's ideas into what some call the "fast and frugal heuristics" school (e.g., Gigerenzer & Selten, 2001). Contrary to other theorists' interpretations of Simon's ideas as suggesting that human judgments and decisions are suboptimal, Gigerenzer and associates argued that many of the heuristics people employ do not merely economize on time and cognitive load. They also perform as well as or better than normative strategies by exploiting structure in real environments (e.g., Gigerenzer et al. 1999). Some heuristics that are fallacies in the casino are effective in real although uncertain environments (e.g., the gambler's fallacy, as shown in Smithson 1997).

Finally, other researchers have investigated the possibility that people think and act as if there are kinds of uncertainty distinct from probability. A classic paper by Ellsberg (1961) presented experimental demonstrations that people are influenced in allegedly counter-normative ways by ambiguity, the extent to which a probability is imprecisely specified. He took the position that the norms proposed by Bayesians were inadequate, and that it is reasonable to be influenced by imprecision.

His findings stimulated extensive research and debate over whether "ambiguity aversion" is irrational or not (see Camerer & Weber 1992 for a review). Smithson (1999) reported experiments showing that people also are influenced by uncertainty arising from conflicting information from equally credible sources. Cabantous (2007) replicated and extended Smithson's findings on a sample of professional insurers. These efforts in psychology and behavioral economics have been paralleled by the creation of alternative formal uncertainty frameworks such as fuzzy logic (Zadeh 1965), belief functions (Shafer 1976), and imprecise probability theories (e.g., Borel 1943/1962, Kyburg 1961, Smith 1961, and Walley 1991). The normative status of these frameworks and whether they are reducible to probability after all are vigorously debated.

Probability Weighting Functions

There is a large body of empirical and theoretical work on subjective probability judgments that considers them in terms of weighting functions. A widely accepted account may be summarized in part by saying that people tend to overweight small and underweight large probabilities (Kahneman & Tversky, 1979), although this account nevertheless is disputed. Note that this claim does not imply that people are under- or over-estimating the probabilities, but instead differentially weighting them when using them for decisions. Rank-dependent expected utility theory (e.g., Quiggin, 1993) reconfigures the notion of a probability weighting function by applying it to a cumulative distribution whose ordering is determined by outcome preferences. Cumulative prospect theory (Tversky & Kahneman, 1992) posits separate weighting functions for gains and losses, on grounds that people are loss-averse in the sense that they are more pessimistic about probabilities of losses and weigh losses more heavily than gains.

Two psychological influences have been offered to explain the properties of probability weighting functions. The hypothesis proposed in Tversky & Kahneman (1992) is “diminishing sensitivity” to changes that occur further away from the reference-points of 0 and 1. This would account for the inverse-S shape, or curvature, of the weighting function typically found in empirical studies (e.g., Camerer & Ho, 1994). For instance, a change from .05 to .10 is seen as more significant than a change from .30 to .35, but a change from .65 to .70 is viewed as less significant than a change from .90 to .95.

Gonzales and Wu (1999) added the notion of “attractiveness” of a gamble to account for the elevation of the weighting curve. The magnitude of consequences affects both the location of the inflection-point of the curve and its elevation. Large gains tend to move the inflection point to the left and large losses move it to the right (e.g., Etchart 2004).

Probabilities from Experience vs Description

Until recently the psychology of probability judgments rested chiefly on evidence from judgments made about descriptions of populations of events or event outcomes provided by experimenters. Little attention was paid to judgments made on the basis of experience, with participants sampling events randomly generated from populations with specified event probabilities (not known by the participants) and estimating probabilities on that basis. An experiment reported in Hertwig, Barron, Weber and Erev (2004) indicated that probability judgments based on experience differ importantly from those based on description. Their primary finding was that in contrast to the view that people overestimate the probabilities of rare events, experience-based judgments *underestimate* the probabilities of such events.

Hertwig et al. called for two theories of probability judgment, one each for judgments based on descriptions and those based on experience. They proposed two explanations account for under-weighting the probability of rare events under experience: Small samples and recency effects (i.e., a tendency to give greater weight to the most recent experiences). They noted that unlike humans, other animal species must entirely base their probability judgments on experience. They cited a study of foraging decisions made by bees (Real 1991), which concluded that the bees under-weight rare instances of food and over-weight common instances because, among other factors, bees' samples from foraging experiences are truncated due to memory constraints.

However, Fox and Hadar (2006) pointed out that Hertwig et al.'s participants showed little or no underweighting of rare events when their judgments were compared to the actual sample probabilities (i.e., the relative frequencies obtained in the participant's sample) instead of the parent population probabilities. They concluded that the apparent underweighting phenomenon is explicable entirely in terms of sampling error and not any psychological influences, a conclusion subsequently supported by others (e.g., Rakow, Demes, & Newell, 2008). Researchers responded by attempting to alter the experienced samples so that they more closely match the population probabilities. One effective approach has been to describe the samples seen by each participant in an experience condition to a yoked ‘‘partner’’ in the description condition (Rakow et al., 2008). In a second approach Camilleri and Newell (2011) implemented an algorithm providing small corrections to the sequence of observations to bring the observed payoffs more in line with their objective counterparts. With both methods the description-experience gap largely disappears. Nonetheless, the debate continues over whether this gap is purely due to bias from sampling variability. Rakow and Newell (2010) have recently urged researchers to study the unique contributions that description and experience make, *in combination*, to probability judgments and decisions based on them. Most recently Camilleri and Newell (2011: 383) have argued that "repeated and consequential choice... appears to be the crucial element for underweighting to occur in the absence of sampling bias."

Probability Judgment Heuristics

Numerous heuristics have been hypothesized to explain people’s probability judgments, especially those deviating from the prescriptions of probability theories. We shall review four of these. The question of what could explain biases in probability judgments was first addressed systematically by Kahneman and Tversky (Kahneman and

Tversky 1972, 1973; Tversky & Kahneman 1974). They proposed three judgment heuristics to account for biases: anchoring-and-adjustment, representativeness and availability.

The anchoring-and-adjustment heuristic involves two claims. First, people are suggestible to arbitrary numerical estimates when engaged in a numerical estimation task. In an experiment demonstrating anchoring, a wheel yielding a number between 1 and 100 was spun in front of participants, who were then asked to estimate what percentage of U.N. countries were in Africa. The median estimate for a group whose starting number was 10 was that 25 percent of U.N. countries were African; the median estimate for the group whose starting number was 65 was nearly twice that: 45 percent. The wheel number biased their estimates even though they knew that the number was irrelevant (Tversky & Kahneman 1974).

The second claim is that people insufficiently adjust away from an initial estimate when presented with evidence. A version of this claim specific to probability judgments is older than the anchoring and adjustment hypothesis (Phillips & Edwards 1966) and is sometimes called ‘conservatism,’ i.e., when the difference between people’s prior and posterior probability estimates is less than that prescribed by Bayes' theorem. The mechanisms that underpin anchoring and adjustment have been debated (e.g., Epley & Gilovich, 2004). Initially, the bias towards initial anchors was ascribed to insufficient adjustment as in conservatism, but later research suggested alternative influences such as confirmation bias.

The representativeness heuristic is that people judge probability via similarity or prototypicality. That is, they ask themselves how similar an item is to a typical member of the category in order to assess the probability that the item belongs to that category. This heuristic has been used to explain people's neglect of relevant base-rates when judging single-event probabilities, and tendency to produce estimates too far from the mean when making predictions based on imperfect predictors. Kahneman and Tversky (1973) demonstrated this heuristic by showing that one group of subjects' judgments of how likely a student was to be majoring in each of nine fields closely matched the judgments by another group of how similar that student was to the typical student in each of those fields.

The availability heuristic is that people judge the probability of an event by how readily previous occurrences come to mind. It is easier to think of English words beginning with K than words that have K as the third letter so most people judge that words of the first kind are the more numerous, when in fact the latter is more numerous. The availability heuristic has been used to explain the Catch-All Underestimation Bias (CAUB, in Fischhoff, Slovic & Lichtenstein 1978), whereby if event categories are combined under a single super-set then the probability people assign to the super-set typically is less than the sum of the probabilities people assign to the component categories (see also Tversky & Koehler 1994).

A more recent example of a probability judgment heuristic is the affect heuristic, a tendency to assess probabilities of outcomes based on how one feels about those outcomes. According to Slovic and Peters (2006), people judge an outcome as less risky if they are favorably disposed towards it, and they consider it more risky if their feelings about it are negative. For example, people fear radiation from nuclear power plants more than they fear radiation from medical X-rays, whereas for most people X-rays pose the greater threat.

These heuristics have stimulated considerable research but also serious criticism. For example, Gigerenzer (1996: 592) averred that the representativeness, availability, and anchoring heuristics are unfalsifiable "… because, post hoc, one of them can be fitted to almost any experimental result. For example, base-rate neglect is commonly attributed to representativeness. However, the opposite result, overweighting of base rates (conservatism), is as easily 'explained' by saying the process is anchoring…" Even a more sympathetic author such as Bar-Hillel (1984) admitted that the counter-normative neglect of three factors (base-rate, sample size, and prediction error) in probability judgments is distinct from the representativeness heuristic and not necessarily explained by it. Attempts to develop more elaborated models of the cognitive processes underlying probability judgments have met with mixed success (see below).

Partition Dependence and Additivity

The aforementioned CAUB phenomenon has been explained not only by the availability heuristic, but also by partition dependence. On grounds of insufficient reason, a probability of $1/K$ is assigned to K mutually exclusive possible events when nothing is known about the likelihood of those events. Fox and Rottenstreich (2003) presented evidence that subjective probability judgments are typically biased towards this ignorance prior. Thus, probability judgments influenced by the ignorance prior are partition dependent. Fox and Clemen (2005) found evidence that this dependence decreases as domain knowledge increases, but that even experts in decision analysis are susceptible to it.

Partition dependence poses problems for probability assignments in two respects. First, it may be unjustified because there is a normatively correct partition. For instance, Fox and Rottenstreich (2003) posed the question of how likely Sunday is to be the hottest day of the week. The principle of insufficient reason would suggest that 1/7 is the correct answer, so their demonstration that people can be induced to partition the events into just two possibilities (Sunday is or is not the hottest day) and therefore assign a probability of 1/2 indicates that those people are anchoring onto an incorrect partition.

The second difficulty arises when there is no normatively correct partition or the sample space is ambiguous. Consider a bag containing 10,000 marbles whose colors are completely unknown to us. How should we use the principle of insufficient reason to judge the probability of drawing a red marble from this bag? Also, consider a scenario in which we are told that Chris, James and Pat are applying to an engineering firm and are then asked to estimate the probability that each of them is hired by the firm. It is unclear whether there is only one position available or multiple positions, or whether these three are the only applicants. Thus, equally defensible partitions could yield ignorance priors of 1/3 each, 1/2 each, or p each, where p is any rational number such that $0 < p \leq 1$. The assignment $p = r/k$, for $k \geq r$, is consistent with assuming there are r positions and k applicants.

Smithson et al. (2011) used this scenario in experiments to show that a simple cue asking people to nominate which candidate has the highest probability of being hired induces more people to constrain their probabilities to strict additivity (i.e., summing to 1 and thus anchoring on a 3-fold partition). Moreover, they found that Japanese respondents were less likely to insist on additivity than Australians. Gelfand and Christakopoulou (1999) report that Americans are more likely to assume a fixed pie (zero-sum) in negotiations than Greeks, which they attribute to the individualism in the former and collectivism in the latter culture. Smithson et al. (2011) found indirect evidence that this could account for the Japanese-Australian difference.

A striking example of strict additivity where none is required comes from a study by Sopena (2009). Australian participants were presented with descriptions of migrants from Syria and Canada and were asked to make judgments regarding the degree to which they considered these targets as prototypically Australian and non-Australian, on a rating scale from 0 to 20 for each judgment. Targets were described as either highly threatening or non-threatening on a variety of issues (e.g., fundamentalist religious orientation). Degrees of membership in the sets of Australians and Syrians or Canadians need not sum to 20 (the sets can overlap). However, Smithson's reanalysis of Sopena's data revealed that respondents' ratings summed exactly to 20 more often for high-threat and Syrian targets. These findings suggest that additivity of group membership increases under perceived threat or when the target is an outgroup member, despite the absence of any normative justification for additivity.

Another influence is a predisposing cognitive bias or stereotype that bolsters a belief that a resource is zero-sum. In Meegan's (2010) experiments participants were undergraduates at a university that does not "grade on the curve," but instead awards grades on the basis of fixed criteria rather than fixed quotas. Thus, grades at that university are not zero-sum. Nevertheless, when participants were asked to predict the grade of a student after they had been shown a skewed distribution of grades already assigned in the same class where the majority were high grades, they predicted a lower grade than participants who had been shown a symmetric distribution. However, the effect disappeared when participants viewed a grade distribution where most grades were low. Meegan concluded that for desirable outcomes people tend towards zero-sum thinking when presented with others' gains but not when presented with their losses.

Calibration and Overconfidence

Even if human probability judgments are not accurate, they may still be well-calibrated in the sense that they are neither consistently too high nor too low. In a classic study, Murphy and Winkler (1977) examined the calibration of 25,000 weather forecasts made over a four-year period which included probability estimates of rain, snow, and other weather events and established that their calibration was very good. However, a review by Lichtenstein, Fischhoff, and Phillips (1982) of the empirical literature on lay confidence judgments indicated that people tend to be over-confident when their confidence is high and under-confident when their confidence is low.

Likewise, a large empirical literature on subjective confidence-interval estimation tasks suggests that people are badly calibrated (overconfident) in the sense that they construct intervals that are too narrow for the confidence level nominated (e.g., Alpert & Raiffa, 1982; Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S., 1999). Nor is this confined to laypeople. In a study of experts' judgmental estimates (Russo & Schoemaker, 1992) in which business managers estimated 90% confidence-intervals for uncertain quantities in their areas of expertise (e.g. petroleum, banking, etc), the hit rates obtained in various samples of managers ranged from 38% to 58%. These are performance levels similar to those typically found in studies of lay people, which indicates that domain expertise does not necessarily confer calibration when it comes to subjective probability estimation.

Yaniv and Foster (1995, 1997) suggested that judgments and evaluations of subjective interval estimates are the product of two competing objectives: accuracy and informativeness. They hypothesized that the patterns of preference ranking for judgments support a simple trade-off model between precision (width) of interval estimates and absolute error which they characterized by the error-to-precision ratio. Both papers presented arguments and evidence that people tend to prefer narrow but inaccurate interval estimates over wide but accurate ones, i.e., they value informativeness more than accuracy. For instance, study 3 in Yaniv and Foster (1995) asked participants to choose between two estimates, (A) [140,150] and (B) [50,300]. They were told that the correct answer was 159. A large majority (90%) of the respondents preferred estimate A over B, although only the latter interval includes the correct answer.

Some researchers found that the format for eliciting interval estimates influences overconfidence. Soll and Klayman (2004) compared overconfidence in interval estimates using three elicitation methods: Range, two-point intervals and three-point intervals. The range method simply asks for, say, a 90% subjective confidence interval. The two-point method asks for a lower limit with a 95% chance of falling below the true value and then an upper limit with a 95% chance of falling above it. The three-point method adds a ``midpoint'' estimate with a 50% criterion. They found the least overconfidence for the three-point method, and the greatest for the range method. Their explanation was that the two-and three-point methods encourage people to sample their knowledge at least twice, whereas the range method is treated by most people as a single sample from their knowledge base. Other aspects of the task that have been investigated include the extremity of the confidence criterion and the nature of the scale used for elicitation (e.g., the study of graininess effects in Yaniv & Foster 1997). For example, Garthwaite & O'Hagan (2000) found that tertiles —the 1/3 and 2/3 quantiles—yielded better calibration than more extreme confidence levels.

A major puzzle in this area was the repeated finding that while people are overconfident when they construct intervals, they are reasonably well-calibrated when asked to assign probabilities to two-alternative questions with the same estimation targets (Klayman, et al. 1999). An example of such a task is asking the respondent whether the population of Thailand exceeds 25 million (yes or no) and then asking for the subjective probability that her answer is correct.

A breakthrough came when Winman, Hansson, and Juslin (2004) revised the two-alternative question format to probability judgments about interval estimates provided to the respondent (e.g., estimating the probability that Thailand's population is between 25 million and 35 million). Comparing these judgments with the intervals elicited from respondents with a fixed confidence criterion (e.g., a 90% subjective confidence interval for Thailand's population), they found that overconfidence was nearly absent in the intervals provided but, as always, high in the elicited intervals. These findings have since been replicated in most, but not all, comparison experiments of this kind (O'Hagan, et al. 2006).

Juslin, Winman, and Hansson (2007) partly accounted for these and related findings by noting that while a sample proportion is an unbiased estimate of the true probability, the sample confidence interval coverage-rate is upwardly biased. They hypothesized that people, as “naïve intuitive statisticians,” are relatively accurate in sampling their own knowledge but treat all sample estimates as if they are unbiased. The implication is that subjective probability judgments of intervals provided to respondents are better-calibrated than intervals produced by respondents to match a coverage-rate.

Conjunction and Conditional Probability Fallacies

Judgments of compound-event and conditional probabilities have been studied extensively. For compound events, most attention has been focused on the so-called *conjunction fallacy*, which is the tendency to violate the conjunction rule that $P(A\&B) \leq \min(P(A), P(B))$. Evidence for these violations first was reported by Kahneman, Slovic and Tversky (1982) and Tversky and Kahneman (1983). Numerous replications followed, and although some scholars questioned whether these really constituted a fallacy (e.g., Wolford, Taylor & Beck 1990), the general consensus has been that the conjunction fallacy is a consistent bias in human reasoning about probabilities.

Nevertheless, a somewhat ironic comparison can be made with Osherson and Smith’s (1981) critique of fuzzy set theory’s adherence to the rule that the degree of membership in the conjunction of two categories cannot exceed membership in either component. They presented counter-examples such as a guppy which is more prototypical of “pet fish” than either of “pet” or of “fish”. Apparently, what is a fallacy for intuitive probability judgments is not for judgments of membership; and whereas mathematics overrules human intuition in probability, human intuition overrules mathematics in categorization.

Several competing explanations for the conjunction fallacy have been advanced. The earliest explanation linked it to the representativeness heuristic (Tversky & Kahneman 1983). In their famous example, they argued that when people judge the probability that Linda is a “bank teller” or “feminist bank teller”, they judge her description to be more similar to “feminist bank teller” than to “bank teller” and therefore assign a higher probability to the conjunction than to the “bank teller” component.

Another explanation was the signed summation model (Yates & Carlson 1986), in which low-probability events are assigned a negative value on a subjective scale and high-probability events a positive value, the idea being that a conjunctive event is judged by the sum of these signed values so that the sum of negative- and positive-scored events will be greater than the negative score of the first event. A third explanation (Thüring & Jungermann 1990) was that the subjective probability of a conjunction typically over-weights the smaller component so that the fallacy arises only when the two component probabilities differ substantially. A fourth account arose from support theory (Tversky & Koehler 1994), which predicts that because probabilities are judged by the number of supporting events, a conjunction will be judged to have greater probability than either of its components.

Agnoli and Krantz (1989) were the first to suggest that competing heuristics might be involved in judgments of conjunctive probabilities and that their dominance would be context-sensitive. Fisk and Pidgeon (1996, 1997) added to this proposal the notion of process-based reasoning, whereby people estimate these probabilities in two stages by anchoring on the smaller component and then adjusting away from that. However, the greatest restraint on the rate at which people commit this fallacy was achieved by Fiedler (1988) who refashioned the problem into judgments about relative frequency. Fiedler asked participants to estimate how many of 100 people "who are like Linda" would fall into the conjunctive and constituent categories. His results were a reduction from 91% to 22% of participants committing the conjunction fallacy.

A related and widely-known fallacy is when people are asked for P(H|E), the probability that an hypothesis H is true given evidence E, when given information about P(E) and the base-rate P(H), they often respond with P(E|H) instead. Numerous examples of this confusion of inverse conditional probabilities have been found in medicine and law, but the most famous example is Tversky and Kahneman's (1982) taxicab problem. Participants were told that 85% of the cabs in a city are Green and 15% Blue; a witness to a hit-and-run accident at night claims the offending cab was Blue; and the witness has been found to identify each of the two colors correctly at nighttime 80% of the time. They were then asked for the probability that the cab in the accident was Blue. Many responded with 80%, i.e., P(E|H).

One implication Tversky and Kahneman drew from responses to the taxicab problem was that participants ignored the base-rate information (i.e., the 15% Blue and 85% Green cabs). A more direct demonstration of *base-rate neglect*, the tendency to ignore or down-weight base-rate information, was Kahneman and Tversky's (1973) study in which participants were told that a brief description of a professional had been randomly drawn from 100 descriptions, 30 engineers and 70 lawyers. When asked the probability that the description was of, say, an engineer their ratings were influenced by how similar they thought the description was to their stereotype of each profession. Even an “information-free” description yielded a probability of .5 from most subjects.

As with the conjunction fallacy, several explanations have been proposed for base-rate neglect. In addition to appeals to representativeness and stereotyping, it has been hypothesized that people consider base-rate information irrelevant (Cohen 1981) or that they confuse P(E|H) with P(H|E) (Eddy 1982). Hamm (1993) found that some subjects’ responses are consistent with the latter explanation and others with the proposal that people interpolate between the base-rate and 1.0.

Also as with the conjunction fallacy, revising the format of the problem into a frequency format greatly increased the percentage of correct answers. Cosmides and Tooby (1990) experimentally demonstrated this with a problem from a study by Casscells, Schoenberger, and Grayboys (1978): If a diagnostic test for a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the probability that a person with a positive test result actually has the disease (if no other information is available)? In the Casscells et al. sample of Harvard Medical School students and staff, only 18% gave the correct answer (.02), and Cosmides and Tooby reported a correct response rate of 12% under the same conditions. Rephrasing the problem in a frequentist way increased that rate to 76% and adding a visual response format (a 10x10 grid with each square representing one person randomly sampled from the population) increased it to 92%.

Communication of Probabilities

In addition to human probability judgments, the communication of probabilistic risk information has been studied extensively (see Budescu & Wallsten 1995 for a thorough review). This research began with the notion that if verbal probability expressions, such as “likely” or “improbable”, have an agreed-upon numerical translation then both elicitation and communication tasks can be simplified by referring to a “dictionary” of these expressions. Indeed, several studies (e.g., Brun & Teign 1988, Erev & Cohen 1990, Wallsten et al. 1993) have reported a widespread preference by people for communicating uncertainties by using verbal expressions rather than numbers. This literature also has debated the question of whether probabilities are better communicated via numbers than by words. There is a long history of attempts to translate verbal probability expressions into numerical form and debates over whether the results are sufficiently reliable and consensual.

In the earliest studies people were asked to nominate single numbers to represent probability expressions (PEs). These studies reported reasonably high intra-subjective reliability (e.g., Lichtenstein & Newman 1967; Beyth-Marom 1982; Budescu & Wallsten 1985; Budescu, Weinberg & Wallsten 1988) and reliable aggregate means (Simpson 1944; Regan, Mosteller & Youtz 1989). However, the same research also revealed considerable inter-subjective variability and overlap among phrases (Stone & Johnson 1959; Lichtenstein & Newman 1967; Beyth-Marom 1982; and Boettcher 1995). Budescu and Wallsten (1985) argued that PEs may lead to ordinal confusion in communication, and Budescu, Weinberg and Wallsten (1988) provided evidence that people vary widely in the PEs they regularly use.

One reasonable interpretation of these findings is that they are symptomatic of vagueness or fuzziness and not just individual differences. Wallsten et al. (1986) established an experimental paradigm in which subjects constructed fuzzy membership functions over the unit interval to translate PEs into numerical terms (see the material on interval evaluation below). Their approach was among the earliest to systematically explore the connection between PEs and *imprecise probabilities*, which include probability intervals and sets of probabilities (see Cozman, this volume). Kent (1964) anticipated this idea by proposing probability intervals as translations of a set of PEs he hoped would be adopted by American intelligence operatives. However, although the British intelligence community eventually adopted this approach, the American intelligence community did not (Kesselman 2008). Translations of PEs into numerical imprecise probabilities seem likely to succeed only in small communities of experts who can agree on nomenclature and, of course, the translation itself.

A recent attempt to impose such a translation on the public at large is the Intergovernmental Panel on Climate Change fourth report (IPCC 2007), which presented a collection of nested intervals corresponding with a set of PEs used throughout their report. Budescu et al. (2009) found that people's estimates of the probabilities corresponding to the PEs in IPCC report sentences were more regressive (towards the middle of the unit interval) than intended by the IPCC authors, an effect that was only partially reduced by embedding an explicit numerical interval in each sentence.

Interpretations of PEs also have proven to be context-dependent. Weber and Hilton (1990) investigated outcome severity in the context of medical scenarios and found that PEs were mapped to higher probabilities when associated with more severe events.

Likewise, PEs appear to be vulnerable to partition priming effects (Tsao & Wallsten 1994). A related literature discusses partition effects for imprecise probabilities. Walley (1991) argued that while partition dependence poses an unavoidable problem for likelihood judgments that yield a single probability, imprecise probability judgments need not depend on the state-space partition. For instance, under complete ignorance Walley recommended that we assign a lower probability of 0 and an upper probability of 1 to every event, no matter what the partition is. Whether or when imprecise probability judgments and PEs are or should be partition-dependent is an open question. Smithson and Segale (2009) compared elicited probability intervals with elicited precise probabilities and found that the location of the intervals were as influenced by partition priming as precise probabilities. However, they also found that some respondents constructed wider intervals when primed with an incorrect partition, indicating that respondents still bore the correct partition in mind.

Some researchers have highlighted relevant differences between meanings or usages in natural vs. formal language. For instance, negation was found to be asymmetric in its effects, so that “unlikely” is not subjectively equivalent to 1 – “likely”. More specifically, PEs have been found to be more inherently “directional” than numbers (Teigen & Brun 1995) and positive PEs tend to be applied to a wider range of numerical probabilities and outcomes than negative PEs. Smithson et al. (in press) reported more regressive responses and less inter-subjective consensus for negative than positive PEs when respondents were asked to translate them from sentences used in the IPCC fourth report (2007). Outcome valence also may affect the interpretation of PEs, although this issue has not been extensively investigated. Mullet and Rivet (1991) compared probabilities assigned to 24 French expressions used in predictions of children's chances of passing or failing a test. On average, a positive context induced higher probabilities for a given phrase.

All told, a clear implication of this line of research is that PEs are not an effective way to communicate probabilities and should be avoided. However, several studies have found that people are no less Bayesian (Rapoport et al. 1990), no more over-confident (Wallsten et al. 1993), and no worse at betting, bidding and decision making (Budescu & Wallsten 1990, Gonzalez-Vallejo et al. 1994) when they use PEs than when they use numbers. Wallsten et al. (1988) proposed a potential resolution of this apparent contradiction by hypothesizing that for decisional purposes people “resolve” the vagueness in a PE by focusing on a numerical probability that they consider prototypical of the PE. Thus, the case against PEs is not entirely convincing.

Finally, relatively little research has been conducted on motivational influences on the communication of uncertainty, despite the fact that normative concerns about bluffing and bid-ask spreads (the difference between the buyer’s highest price and the seller’s lowest price) extend back to at least Borel (1924/1964: 58) and produced the literature on scoring rules for eliciting “honest” probability judgments (e.g., Brier 1950). Schweitzer and Hsee (2002) experimentally demonstrated that motivational factors exert greater influence on estimations elicited from participants under high uncertainty conditions than under low uncertainty. They argued that greater uncertainty creates leeway for decision makers to justify extreme (and self-serving) claims to themselves. From a normative viewpoint, Seidenfeld, Schervish and Kadane (2011) provided a formal argument that there is no single real-valued scoring rule that can play the same regulatory role for imprecise probabilities as Brier scores can for precise probabilities.

Normative Fundamentalism?

Because so much research on human probability judgments has compared them with some version of a Bayesian framework which is taken as the benchmark of rationality, we shall conclude by briefly noting three lines of criticism that have been leveled at this approach. The more moderate criticism has been that comparisons with Bayesian prescriptions do not sufficiently address questions of how people construct their judgments. Elqayam and Evans (2011) claim that an “ought-is” distinction biases research by restricting attention to normative correlates and neglecting philosophically significant questions that lack a clear standard for normative judgment. Earlier, Gigerenzer (1993) called for greater attention to the mental models and cognitive processes involved in probability judgments and less of a focus on errors and biases. Indeed, recent trends in the literature on probability judgment have been towards greater emphasis on modeling cognitive processes.

A more radical line of criticism has been that the normative standards are inappropriate or miss-specified. For instance, Gigerenzer (1991) and Teigen (1994), among others, have argued that distinctions such as the one between single-event and relative frequency probabilities yield important disputes about normative standards for probability judgments. Gigerenzer points to prominent proponents of the frequentist school of probability theory, who regard the concept of a probability of a unique event as meaningless. He also advocates comparing “apples with apples,” for instance, comparing objective relative frequencies with people's estimates of relative frequencies instead of with their confidence judgments. Likewise, Teigen claims that at least some of the "deviations" from Bayesian prescriptions can be explained by defensible reasoning depending on whether probabilities are being judged on the basis of base-rates (relative frequencies), internal mental states, dispositions, or degrees of plausibility. Crupi, Fitelson, and Temtori (2008) argue that experimentally observed fallacious probability judgments in conjunction problems may be guided by sound assessments of confirmation relations (as in Bayesian confirmation theory).

The third critique has been alluded to earlier, namely that there may be other kinds of uncertainty such as ambiguity or conflict, and when people are influenced by these they are not behaving irrationally. Arguments defending ambiguity aversion include aversion to missing but obtainable information (Ritov & Baron 1990) and sensitivity to variability of outcomes (e.g., Rode et al. 1999). Outcome variance relative to outcome magnitude is an effective predictor of responses to risk in human and non-human species (Weber, Shafir & Blais 2004). This finding has been defended via optimal foraging theory (Caraco 1981), whereby animals choose the option with lower variance if the mean caloric payoff exceeds current need but choose the higher-variance option if the mean payoff falls below current need. Similar arguments may be extended to conflict aversion, along with the question of trustworthiness of the conflicting message sources when they are supposedly based on the same information (Smithson 1999).

These critiques and the research and theory development stemming from them have deepened our understanding of the psychological aspects of probability judgment, and judgments of risk in general. They also have stimulated and contributed to the ongoing debates about the normative status of precise probabilities and candidates for other kinds of uncertainty.

References

Agnoli, F. and Krantz, D.H. (1989), ‘Suppressing Natural Heuristics by Formal Instruction. The Case of the Conjunction Fallacy’, in *Cognitive Psychology*, 21: 515-550.

Alpert, M. & Raiffa, H. (1982), ‘A Progress Report on the Training of Probability Assessors’, in D. Kahneman, P. Slovic, & A. Tversky (eds.) *Judgment under Uncertainty: Heuristics and Biases*, (New York: Cambridge University Press), 294-305.

Bar-Hillel, M. (1984), ‘Representativeness and Fallacies of Probability Judgment,’ in *Acta Psychologica*, 55: 91-107.

Bell, C.R. (1979), ‘Psychological Aspects of Probability and Uncertainty’, in C.R. Bell (ed.), *Uncertain Outcomes* (Lancaster: MTP Press), 5-21.

Beyth-Marom, R. (1982), ‘How Probable is Probable? A Numerical Translation of Verbal Probability Expressions’, in *Journal of Forecasting*, 1: 257-269.

Boettcher, W.A. (1995), ‘Context, Methods, Numbers, and Words: Prospect Theory in International Relations’, in *Journal of Conflict Resolution*, 39: 561-583.

Borel, E. (1921), ‘La Théorie du Jeu et les Équations Intégrales a Noyau Symétrique’, in *Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences*, 173: 1304-1308 (English translation by L.J. Savage (1953) in *Econometrica*, 21: 97-100).

Borel, E. (1924/1964) (Tr. H.E. Smokler), ‘Apropos of a Treatise on Probability’, in H.E, Kyburg & H.E. Smokler (eds.), *Studies in Subjective Probability* (New York: Wiley), 45-60.

Borel, E. (1943/1962), *Probabilities and Life*, (New York:Dover), French Original *Les Probabilités et la Vie*, Presses Universitaires de France.

Brier, G. W. (1950), ‘Verification of Forecasts Expressed in Terms of Probability’, in *Monthly Weather Review,* 78: 1–3.

Brun, W. & Teigen, K.H. (1988), ‘Verbal Probabilities: Ambiguous, Context-Dependent, or Both?’ in *Organizational Behavior and Human Decision Processes*, 41: 390-404.

Budescu, D.V., Broomell, S. & Por, H.H. (2009), ‘Improving Communication of Uncertainty in the Reports of the Intergovernmental Panel on Climate Change’, in *Psychological Science*, 20: 299-308.

Budescu, D.V. & Wallsten, T.S. (1985), ‘Consistency in Interpretation of Probabilistic Phrases’, in *Organizational Behavior and Human Decision Processes,* 36: 391-405.

Budescu, D.V. & Wallsten, T.S. (1990), ‘Dyadic Decisions with Verbal and Numerical Probabilities’, in *Organizational Behavior and Human Decision Processes*, 46: 240-263.

Budescu, D.V. & Wallsten, T.S. (1995), 'Processing Linguistic Probabilities: General Principles and Empirical Evidence', in J. Busemeyer, R. Hastie, D. L. Medin, (eds.), *Decision Making from a Cognitive Perspective,* (San Diego: Academic Press), 275–318.

Budescu , D.V., Weinberg, S. & Wallsten, T.S. (1988), 'Decisions Based on Numerically and Verbally Expressed Uncertainties', in *Journal of Experimental Psychology: Human Perception and Performance*, 14: 281-294.

Cabantous, L. (2007), 'Ambiguity Aversion in the Field of Insurance: Insurers' Attitude to Imprecise and Conflicting Probability Estimates', in *Theory and Decision*, 62: 2219–240.

Camerer, C.F. & Ho, T.H. (1994), 'Violations of the Betweenness Axiom and Nonlinearity in Probability', in *Journal of Risk and Uncertainty*, 8: 167-196.

Camerer, C. & Weber, M. (1992), 'Recent Developments in Modeling Preferences: Uncertainty and Ambiguity', in *Journal of Risk and Uncertainty*, 5: 325-370.

Camilleri, A.R. & Newell, B.R. (2011), 'When and Why Rare Events are Underweighted: A Direct Comparison of the Sampling, Partial Feedback, Full Feedback and Description Choice Paradigms', in *Psychonomic Bulletin & Review*, 18: 377-384.

Caraco, T. (1981), 'Energy Budget, Risk, and Foraging Preferences in Dark-Eyed Juncos', in *Behavioral Ecological Sociobiology*, 8: 820-830.

Cohen, J. (1954), 'Conjecture and Risk', in *The Advancement of Science Reports of the British Association*, 11: 333-339.

Cohen, J. (1960), *Chance, Skill and Luck*, (Harmondsworth: Penguin).

Cohen, J. (1964), *Behaviour in Uncertainty*, (London: Allen & Unwin).

Cozman, F.

Crupi, Fitelson, and Temtori (2008), Probability, confirmation and the conjunction fallacy, in *Thinking and Reasoning*, 14: 182-199.

Comment [MS1]: Chapter reference to go here.

de Finetti, B. (Tr. H.E. Kyburg) (1937/1964), 'Foresight: Its logical laws, its subjective sources', in H.E, Kyburg & H.E. Smokler (eds.), *Studies in Subjective Probability* (New York: Wiley), 93-158.

Edwards, W. (1954), 'The Theory of Decision Making', in *Psychological Review*, 51: 380-417.

Edwards, W. (1961), 'Behavioral Decision Theory', in *Annual Review of Psychology*, 12: 473-498.

Elqayam, S. & Evans, J.St. B.T. (2011), 'Subtracting "Ought" from "Is": Descriptivism versus Normativism in the Study of Human Thinking', in *Behavioral and Brain Sciences*, 34: 233 –290.

Ellsberg, D. (1961), 'Risk, Ambiguity, and the Savage Axioms', in *Quarterly Journal of Economics*, 75: 643-669.

Epley, N., & Gilovich, T. (2004), 'Are Adjustments Insufficient?' in *Personality and Social Psychology Bulletin*, 30: 447-460.

Erev, I. and Cohen, B.L. (1990), 'Verbal versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox', in *Organizational Behavior and Human Decision Processes*, 45: 1-18.

Etchart-Vincent, N. (2004), 'Is Probability Weighting Sensitive to the Magnitude of Consequences? An Experimental Investigation on Losses', in *Journal of Risk and Uncertainty*, 28: 217-235.

Fiedler, K., (1988), 'The Dependence of the Conjunction Fallacy on Subtle Linguistic Factors', in *Psychological Research*, 50: 123-129.

Fischhoff, B., Slovic, P. and Lichtenstein, S. (1978), 'Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation', in *Journal of Experimental Psychology: Human Perception Performance* 4: 330–44.

Fisk, J.E. and Pidgeon, N.F. (1996), 'Component Probabilities and the Conjunction Fallacy: Resolving Signed Summation and the Low Probability Model in a Contingent Approach', in *Acta Psychologica*, 94: 1-20.

Fisk, J.E. and Pidgeon, N.F. (1997), 'The Conjunction Fallacy: the Case for the Existence of Competing Heuristic Strategies', in *British Journal of Psychology*, 88: 1-27.

Fox, C.R. & Clemen (2005), 'Subjective Probability Assessment in Decision Analysis: Partition Dependence and Bias toward the Ignorance Prior', in *Management Science*, 51: 1417-1432.

Fox, C.R. & Rottenstreich, Y. (2003), 'Partition Priming in Judgment under Uncertainty', in *Psychological Science*, 13: 195-200.

Garthwaite, P. H. and O'Hagan, A. (2000), 'Quantifying expert opinion in the UK water industry: An experimental study', in *The Statistician*, 49: 455-477.

Gelfand, M.J. and Christakopoulou, S. (1999), 'Culture and Negotiator Cognition: Judgment Accuracy and Negotiation Processes in Individualistic and Collectivistic Cultures', in *Organizational Behavior and Human Decision Processes*, 79: 248-269.

Gigerenzer, G. (1993), 'The Bounded Rationality of Probabilistic Mental Models', in K.I. Manktelow and D.E. Over (eds.), *Rationality: Psychological and Philosophical Perspectives*, (London: Routledge), 285-313.

Gigerenzer, G. (1996), 'On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky (1996)', in *Psychological Review*, 103: 592-596.

Gigerenzer, G., Todd, P.M., and the ABC Research Group (1999), *Simple Heuristics that Make Us Smart*, (London: Oxford University Press).

Gigerenzer, G. & Selten, R. (eds.) (2001), *Bounded Rationality: The Adaptive Toolbox*, (Cambridge MA: MIT Press).

Gonzales, R. & Wu, G. (1999), ‘On the Shape of the Probability Weighting Function’, in *Cognitive Psychology*, 8: 129-166.

Gonzalez-Vallejo, C.C., Erev, I. & Wallsten, T.S. (1994), 'Do Decision Quality and Preference Order Depend on Whether Probabilities are Verbal or Numerical?' in *American Journal of Psychology*, 107: 157-172.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004), ‘Decisions from Experience and the Effect of Rare Events in Risky Choice’, in *Psychological Science*, 15: 534–539.

Hogarth, R.M. (1980), *Judgment and Choice: The Psychology of Decision*, (Chichester: Wiley).

Juslin, P., Winman, A., and Hansson, P. (2007), ‘The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals’, in *Psychological Review*, 114: 678-703.

Kahneman, D. & Tversky, A. (1972), ‘Subjective Probability: A Judgment of Representativeness’, in *Cognitive Psychology*, 3: 430-454.

Kahneman, D. & Tversky, A. (1973), ‘On the Psychology of Prediction’, in *Psychological Review*, 80: 237-251.

Kahneman, D. & Tversky, A., (1979), ‘Prospect theory: An Analysis of Decision under Risk’, in *Econometrica*, 47: 263-291.

Kahneman, D., Slovic, P. & Tversky, A. (eds.), (1982), *Judgment under Uncertainty: Heuristics and Biases*, (New York: Cambridge University Press).

Kent, S. (1964), 'Words of Estimative Probability', *Studies in Intelligence*, 8: 49-65.

Kesselman, R.F. (2008), *Verbal Probability Expressions in National Intelligence Estimates: A Comprehensive Analysis of Trends from the Fifties through Post 9/11*, (Unpublished Masters Thesis, Mercyhurst College, PA).

Keynes, J.M. (1921), *A Treatise on Probability*, (London: MacMillan).

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S., (1999), 'Overconfidence: It Depends on How, What, and Whom You Ask', in *Organizational Behavior and Human Decision Processes*, 79: 216-247.

Kyburg, H.E. Jr. (1961), *Probability and the Logic of Rational Belief*, (Middletown: Wesleyan University Press).

Lichtenstein, S., Fischhoff, B, & Phillips, B. (1982), 'Calibration of Probabilities: The State of the Art to 1980', in D. Kahneman, P. Slovic, & A. Tversky (eds.) *Judgment under Uncertainty: Heuristics and Biases*, (New York: Cambridge University Press), 306-334.

Lichtenstein, S. and Newman, J.R. (1967), 'Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities', in *Psychonomic Science*, 9: 563-564.

Meegan, D.V. (2010), 'Zero-Sum Bias: Perceived Competition Despite Unlimited Resources', in *Frontiers in Psychology: Cognition*, 1: 1-7.

Mullet, E. and Rivet, I. (1991), 'Comprehension of Verbal Probability Expressions in Children and Adolescents', in *Language and Communication*, 11: 217-22

Murphy, A.H. & Winkler, R.L. (1977), 'Can Weather Forecasters Formulate Reliable Probability Forecasts of Precipitation and Temperature? In *National Weather Digest*, 2: 2-9.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006), *Uncertain Judgements: Eliciting Experts' Probabilities*, (Chichester: Wiley).

Osherson, D.N. & Smith, E.E. (1981), 'On the Adequacy of Prototype Theory as a Theory of Concepts', in *Cognition*, 9: 35-58.

Phillips, L.D. & Edwards, W., (1966), 'Conservatism in a Simple Probability Inference Task,' in *Journal of Experimental Psychology*, 72: 346-354.

Quiggin, J. (1993), *Generalized Expected Utility Theory: The Rank Dependent Model*, (Boston: Kluwer).

Rakow, T. & Newell, B.R. (2010), 'Degrees of Uncertainty: An Overview and Framework for Future Research on Experience-Based Choice', in *Journal of Behavioral Decision Making*, 23: 1–14.

Ramsey, F.P. (1926/1931), 'Truth and Probability', in R.B. Braithwaite (ed.), *The Foundations of Mathematics and Other Logical Essays* (London: Routledge and Kegan Paul), 156-198.

Reagan, R.T., Mosteller, F. and Youtz, C. (1989), 'Quantitative Meanings of Verbal Probability Expressions', in *Journal of Applied Psychology*, 74: 433-442.

Ritov, I. & Baron, J. (1990), 'Reluctance to Vaccinate: Omission Bias and Ambiguity', in *Journal of Risk and Uncertainty*, 5: 49-61.

Rode, C., Cosmides, L., Hell, W. & Tooby, J. (1999), 'When and Why Do People Avoid Unknown Probabilities in Decisions under Uncertainty? Testing Some Predictions from Optimal Foraging Theory', in *Cognition*, 72: 269-304.

Russo, J. E., & Schoemaker, P. J., (1992), 'Managing Overconfidence', in *Sloan Management Review*, 33: 7-17.

Schweizer, M.E. & Hsee, C.K. (2002), 'Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information', in *Journal of Risk and Uncertainty*, 25: 185-201.

Seidenfeld, T. Schervish, M.J. & Kadane, J.B. (2011), 'Forecasting with Imprecise Probabilities', in F. Coolen, G. de Cooman, T. Fetz & M. Oberguggenberger (eds.) *Proceedings of the Seventh International Symposium on Imprecise Probability: Theories and Applications*, Innsbruck, Austria, 317-326.

Shafer, G. (1976), *A Mathematical Theory of Evidence*, (Princeton: Princeton University Press).

Simon, H.A. (1956), 'Rational Choice and the Structure of Environments', in *Psychological Review*, 63: 129-138.

Simon, H. (1982), *Models of Bounded Rationality* (2 vols.), (Cambridge MA: MIT Press).

Simpson, R.H. (1944), 'The Specific Meanings of Certain Terms Indicating Differing Degrees of Frequency', in *Quarterly Journal of Speech*, 30: 328-330.

Smith, C.A.B. (1961), 'Consistency in Statistical Inference and Decision', *Journal for the Royal Statistical Society, Series B*, 23: 1-37.

Smithson, M. (1997), 'Judgment Under Chaos', in *Organizational Behavior and Human Decision Processes* 69: 59-66.

Smithson, M. (1999), ‘Conflict Aversion: Preference for Ambiguity vs. Conflict in Sources and Evidence’, in *Organizational Behavior and Human Decision Processes*, 79: 179-198.

Smithson, M., Budescu, D.V., Broomell, S.B., and Por, H.-H. (in press), ‘Never Say 'Not:' Impact of Negative Wording in Probability Phrases on Imprecise Probability Judgments’, *International Journal of Approximate Reasoning*.

Smithson, M. and Segale, C. (2009), ‘Partition Priming in Judgments of Imprecise Probabilities’, in *Journal of Statistical Theory and Practice*, 3: 169-182.

Smithson, M., Verkuilen, J., Hatori, T. And Gurr, M. (2010), *More than a Mean Difference: New Models and Findings of Partition Priming Effects on Probability Judgments*, Paper presented at the 31st Annual Conference of the Society of Judgment and Decision Making, Nov. 19-22, St Louis, MO.

Soll, J. & Klayman, J. (2004), ‘Overconfidence in Interval Estimates’, in *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30: 299-314.

Sopena, A. (2009), *Somewhere in Between: The Contribution of Ethnicity, Threat and Social Identity to the Production of Marginalizing Racism*, Unpublished Honours Thesis (Canberra: The Australian National University).

Stanovich, K.E., & West, R.F. (2000), ‘Individual Differences in Reasoning: Implications for the Rationality Debate’, in *Behavioral & Brain Sciences*, 23: 645–665.

Stone, D.R. and Johnson, R.J. (1959), ‘A Study of Words Indicating Frequency’, in *Journal of Educational Psychology*, 50: 224-227.

Teigen, K. H. (1994), ‘Variants of Suvjective Probabilities: Concepts, Norms, and Biases’, in G. Wright & P. Ayton (eds.), *Subjective Probability*, (London: Wiley), 211-238.

Thüring, M. & Jungermann, H. (1990), ‘The Conjunction Fallacy: Causality vs. Event Probability’, in *Journal of Behavioral Decision Making*, 3: 61-74.

Tversky, A. and Kahneman, D. (1974), ‘Judgment under Uncertainty: Heuristics and Biases’, in *Science*, 185: 1124-1131.

Tversky, A. & Kahneman, D. (1983), ‘Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment’, in *Psychological Review*, 90: 293-315.

Tversky, A. & Kahneman, D. (1992), ‘Advances in Prospect Theory: Cumulative Representation of Uncertainty’, in *Journal of Risk and Uncertainty*, 5: 297-323.

Tversky, A., and Koehler, D. (1994), ‘Support theory: A nonextensional representation of subjective probability’, in *Psychological Review* 101: 547–67.

von Neumann, J. and Morgenstern, O. (1944), *Theory of Games and Economic Behavior* (Princeton: Princeton University Press).

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, (London: Chapman and Hall).

Wallsten, T.S., Budescu, D.V. & Erev, I. (1988), 'Understanding and Using Linguistic Uncertainties', in *Acta Psychologica*, 68: 39-52.

Wallsten, T.S., Budescu, D.V., Rapoport, A., Zwick, R. & Forsyth, B.H. (1986), ‘Measuring the Vague Meanings of Probability Terms’, in *Journal of Experimental Psychology: General*, 115: 348-365.

Weber, E.U. and Hilton, D.J. (1990), ‘Contextual Effects in the Interpretations of Probability Words: Perceived Base Rate and Severity of Events’, in *Journal of Experimental Psychology: Human Perception and Performance*, 16: 781-789.

Weber, E.U., Shafir, S., & Blais, A-R. (2004), ‘Predicting Risk Sensitivity in Humans and Lower Animals: Risk as Variance or Coefficient of Variation’, in Psychological Review, 111: 430-445.

Winman, A., Hansson, P., and Juslin, P. (2004), ‘Subjective probability intervals: How to cure overconfidence by interval evaluation’, in *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30: 1167-1175.

Wolford, G., Taylor, H.A., & Beck, J.R. (1990), ‘The Conjunction Fallacy?’ in *Memory and Cognition*, 18: 47-53.

Yaniv, I. & Foster, D.P. (1995), ‘Graininess of judgment under uncertainty: An accuracy-informativeness tradeoff’, in *Journal of Experimental Psychology: General*, 124: 424-432.

Yaniv, I. & Foster, D.P. (1997), ‘Precision and accuracy of judgmental estimation’, in *Journal of Behavioral Decision Making*, 10: 21-32.

Yates, J.F. & Carlson, B.W. (1986), ‘Conjunction Errors: Evidence for Multiple Judgment Procedures Including ‘Signed Summation’’, in *Organizational Behavior and Human Decision Processes*, 37: 230-253.

Zadeh, L. (1965), ‘Fuzzy Sets’, in *Information and Control*, 8: 338-353.